JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D23, PAGES 29,135–29,144, DECEMBER 16, 2000

Assimilation of satellite observations of long-lived chemical species in global chemistry transport models

Boris V. Khattatov,[1] Jean-Francois Lamarque,[1] Lawrence V. Lyjak,[1] Richard Menard,[2] Pieternel Levelt,[3] Xuexi Tie,[1] Guy P. Brasseur,[4] and John C. Gille[1]

Abstract. Use of data assimilation techniques such as optimal interpolation or the Kalman filter in global chemistry transport models (CTM) is becoming more common. However, owing to high computational requirements, it is often difficult to apply these techniques to multidimensional models containing extensive photochemical schemes. We present a sequential assimilation approach developed for use with general global chemistry transport models. It allows fast assimilation and mapping of satellite observations and provides estimates of analysis errors. The suggested data assimilation scheme evolved from the one described by *Levelt et al.* [1998]. It is a variant of the suboptimal Kalman filter and is based on ideas described by *Menard et al.* [2000] and *Menard and Chang* [2000]. One of the most important features of the developed scheme is its ability to routinely estimate variance of the analysis and to predict variance evolution in the model. The developed technique (or its variants) has been successfully interfaced with a number of different global models and used for assimilation of several types of measurements, including aerosol extinction ratios. Some of these experiments are described by *Lamarque et al.* [1999] and W. D. Collins et al. (Forecasting aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX, submitted to *Journal of Geophysical Research*, 2000, hereinafter referred to as Collins et al., submitted manuscript, 2000). We illustrate the method using assimilation of ozone observations made by the Upper Atmosphere Research Satellite/Microwave Limb Sounder in the three-dimensional chemistry transport model ROSE [Research for Ozone in the Stratosphere and its Evolution; *Rose and Brasseur*, 1989].

1. Introduction

Several research groups have recently demonstrated that data assimilation techniques can be very successfully used for analysis of atmospheric chemical observations. To a large degree, this success is due to bringing together traditional atmospheric photochemical modeling and sophisticated mathematical methods of estimation theory. Rigorous error analysis methods, such as the Kalman filter, developed in estimation theory are likely to yield significant benefits when applied to challenging problems in atmospheric chemistry.

Lyster et al. [1997] for the first time applied the Kalman filter to assimilation of satellite observations in a global atmospheric transport model. Upper Atmosphere Research Satellite (UARS) observations of a single inert tracer (CH_4) were assimilated in a two-dimensional (in longitude and latitude) transport model at a single potential temperature level. The model contained no chemical processes. Because of the relative simplicity of the employed model, it was possible to use the full Kalman filter for estimation of global error covariances and tracking error covariances evolution in the model. The work of *Lyster et al.* [1997] highlighted the significant computational difficulties arising when applying the full Kalman filter to global multi-dimensional atmospheric models.

Menard et al. [2000] and *Menard and Chang* [2000] studied several approaches to implementing the Kalman filter for assimilating UARS data in the two-dimensional transport model. In particular, they found that the standard Kalman filter covariance propagation is highly inaccurate for this problem. They suggested an alternative implementation based on the continuum error covariance. Second, *Menard et al.* [2000] and *Menard and Chang* [2000] suggested a rather extensive diagnostic process based on the chi-square method for adjusting free parameters in the assimilation scheme. Finally, they proposed a Kalman filter formulation based on the relative errors.

In this work we extend the methodology proposed by *Menard et al.* [2000] and *Menard and Chang* [2000] for use with general three-dimensional chemistry transport models and multiple types and sources of satellite observations. The multidimensional nature of the problem leads to rather extensive computational requirements in terms of both CPU power and memory. Thus particular attention has been paid to the efficient design of the associated computer code, and certain simplifications were made to facilitate the practical implementation. Furthermore, since satellite observations are taken sequentially in time, it is possible to somewhat decrease the size of the problem (in the space of observations) by decreasing the length of the assimilation window. This is because fewer observations will have to be assimilated at once and the computing time required per assimilation procedure growths faster than linear with number of observations. Additionally, smaller assimilation window

[1]National Center for Atmospheric Research, Boulder, Colorado.
[2]Air Quality Research Branch, Meteorological Service of Canada, Dorval, Quebec, Canada.
[3]Royal Dutch Meteorological Institute, De Bilt, Netherlands.
[4]Max Plank Institute for Meteorology, Hamburg, Germany

Paper number 2000JD900466.
0148-0227/00/2000JD900466$09.00

decreases the difference between the time observations were taken and the time they were assimilated.

At the time of this writing, the technique described in this manuscript (or its variants) has been successfully used with a number of different global models (ROSE, MOZART2 (Model for OZone And Related chemical Tracers), MATCH (Model of Atmospheric Transport and CHemistry), TM3 (Tracer Model 3)) for assimilation of several different types of measurements: UARS MLS ozone, MAPS CO, IMG CO, MOPITT CO, GOME ozone, SBUV ozone, and aerosol extinction from AVHRR and PATHFINDER. Some of these experiments are described by *Lamarque et al.* [1999] and Collins et al. (submitted manuscript, 2000).

A characteristic feature of the Kalman filter is computation of the time evolution of the forecast error covariance. While this feature is certainly extremely desirable, it is not necessary for developing and implementing data assimilation methodologies. For instance, *Riishojgaard et al.* [2000] have designed a very successful data assimilation technique that does not include explicit calculations of variance evolution.

In this work we concentrate on the sequential approach to data assimilation. It should be noted that there exist numerous and successful implementations of the variational method to multidimensional atmospheric models [e.g., *Eskes et al.*, 2000; *Elbern et al.*, 1997] Discussion of these methods is beyond the scope of this manuscript.

2. Model

In this work we applied the developed data assimilation technique to the three-dimensional chemistry transport model Research for Ozone in the Stratosphere and its Evolution (ROSE) [*Rose and Brasseur*, 1989] for assimilation of UARS MLS ozone measurements in the stratosphere. The model contains an extensive set of photochemical reactions as well as heterogeneous processes. The photochemical scheme contains approximately 50 species and includes oxygen, nitrogen, carbon, chlorine, hydrogen, and bromine species. The rate constants as well as the absorption cross sections are taken from *DeMore et al.* [1994]. While there exist newer revisions of this publication, previous studies with the model demonstrated that the chemical scheme is adequate for our purposes. Previously, the model has been used for studies of ozone and other trace gases [*Granier and Brasseur*, 1991]. The off-line version of ROSE used in this study is driven by the daily United Kingdom Meterological Office (UKMO) stratospheric analysis data calculated for the UARS mission [*Swinbank and O'Neill*, 1994]. The horizontal resolution of the model is 5° latitude and 11 25° longitude. The model contains 19 layers in the vertical from 316 mbar up to 0.316 mbar. These pressure levels coincide with the standard UARS levels and the UKMO stratospheric analysis levels. The transport scheme is semi-Lagrangian, described by *Smolarkiewicz and Rasch* [1991]. The model chemical time step is 30 min, and the dynamical time step is 2 hours.

3. UARS MLS Observations

The MLS instrument on board UARS [*Reber et al.*, 1993] observes the microwave atmospheric limb emissions on a global scale during both day and night. On a given day, MLS normally collects approximately 1500 profiles. The instrument measures profiles of several trace gases including O_3 with vertical resolution of about 6 km. Throughout this manuscript, we use UARS MLS level 3 AT data.

In version 4 the retrievals are made on alternate standard UARS pressure levels. The retrieved mixing ratios are subsequently interpolated to the standard UARS pressure levels, which coincide with the model vertical levels.

It is believed that the MLS version 4 ozone data used in this study are reliable from about 46 to 0.46 hPa [*Froidevaux et al.*, 1996]. In most cases the accuracy and approximate precision are 0.3-0.4 ppmv and 0.2-0.3 ppmv, respectively, as estimated from the UARS/MLS Data Quality Document (http://mls.jpl.nasa.gov/lucien/daac\document\v4). Only data at the retrieved levels that are considered to be reliable, 46.4, 21 5, 10.0, 4.6, 2.1, 1.0, and 0.464 mbar, were used for the assimilation.

Observational error was computed from the MLS level 3AT precision and accuracy information given in the UARS/MLS Data Quality Document. For each pressure level the relative error was computed as

$$\sigma_r^2 = a^2 + p^2$$

where a and p represent relative accuracy and precision obtained from the UARS/MLS Data Quality Document. Hence the absolute error σ of measurement y is

$$\sigma = \sigma_r y$$

For the seven pressure levels used in the assimilation the computed relative errors σ_r are 0.23, 0.06, 0.05, 0.05, 0.06, 0 14, and 0.25, respectively

4. Assimilation Scheme

Global chemistry transport models produce time series of estimates of concentrations of atmospheric chemicals at nodes of a three-dimensional grid. Let us arrange model estimates of concentration of a particular chemical at time t in vector $\mathbf{x}$ with dimension N_x. Formally, integration of the model $\mathbf{M}$ can be written as

$$\mathbf{x}_{t+\Delta t} = \mathbf{M}(t, \mathbf{x}_t) . \quad (1)$$

Let vector $\mathbf{y}$ contain observations of a quantity somehow related to concentrations of that chemical at approximately the same time. Examples of such observations include measurements of total column amount, partial (e.g., tropospheric) column amount, or constituent concentrations at a few selected pressure levels, as in the case of MLS observations.

Geographical locations of observations usually differ from locations of the model grid. We will assume that the relationship between the observed quantity $\mathbf{y}$ and estimates of concentrations at the geographical location of the observation is linear. Thus the connection between $\mathbf{x}$ and $\mathbf{y}$ can be established through combination of two linear operators, $\mathbf{I}$ and $\mathbf{A}$. The horizontal linear interpolation operator $\mathbf{I}$ represents interpolation of the state variables (constituent concentrations) from the horizontal locations of the model grid points to the locations of the observations. In this work we use a simple bi-linear interpolation. Operator $\mathbf{A}$ denotes the relationship between the observed quantities and estimated constituent concentrations at the geographical locations of observations. It can be as simple as a single row containing all zeros and a single "1," in which case it corresponds to measuring a concentration of a chemical at a single pressure level. It can also be a row containing all ones, in the case when total column content is observed. In certain cases

(e.g., MAPS observations as described by *Lamarque et al.* [1999]) it can be of complicated shape. In the case of MLS observations and the ROSE model, operator **A** is a matrix containing 7 rows (number of observed pressure levels) and 19 columns (number of model pressure levels). Since the observed MLS pressure levels are a subset of the model pressure levels (standard UARS levels), each row of the matrix contains "1" in the position corresponding to the observed level and zeros elsewhere

The connection between **x** and **y** is therefore as follows:

$$\mathbf{y} = \mathbf{A}(\mathbf{I}(\mathbf{x})) .$$

To simplify notation, we define observational operator **H** as

$$\mathbf{H}(\mathbf{x}) = \mathbf{A}(\mathbf{I}(\mathbf{x}))$$

so that

$$\mathbf{y} = \mathbf{H}(\mathbf{x}) \tag{2}$$

The analysis problem is then to find the "best" value of **x**, which inverts (2) for a given **y** allowing for observation errors and other prior information [*Lorenc*, 1986] The solution is usually called the analysis, $\mathbf{x}^a$ Prior information is given by independent estimates of vector **x**, or background and error covariance matrices of the background and observations In most cases, dimensions of vectors **x** and **y** will be different, and this problem will be either overdetermined or underdetermined. Therefore inversion of (2) should be done in the statistical sense

Inversion of (2) is performed at fixed time intervals using all observations collected during this period Such a time interval is commonly called the assimilation window. The model forecast at the beginning of the assimilation window, $\mathbf{x}_t$, is considered to be the a priori background estimate.

It can be shown [e g , *Lorenc*, 1986] that at each time subsequent step the solution of (2) is given by

$$\mathbf{x}_t^a = \mathbf{x}_t + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_t) \tag{3}$$

$$\mathbf{K} = \mathbf{B}_t\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}_t\mathbf{H}^{\mathrm{T}} + \mathbf{O} + \mathbf{R})^{-1} . \tag{4}$$

Here $\mathbf{B}_t$ is the forecast error covariance at time t **O** is the error covariance matrix of the observations and **R** is the representativeness error covariance associated with errors of interpolation and discretization. Matrix **K** is called the Kalman gain matrix

The analysis error covariance is expressed as [*Lorenc*, 1986]:

$$\mathbf{B}_t^a = \mathbf{B}_t - \mathbf{B}_t\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}_t\mathbf{H}^{\mathrm{T}} + \mathbf{O} + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}_t . \tag{5}$$

Once inversion of (2) is performed, the obtained concentrations, $\mathbf{x}_t^a$, can be used as the initial condition for the chemistry transport model **M** to predict constituent concentrations at a later time (beginning of the next assimilation window) according to (1):

$$\mathbf{x}_{t+\Delta t} = \mathbf{M}(t, \mathbf{x}_t^a) . \tag{6}$$

In the extended Kalman filter method, evolution of the error covariance is obtained using linearization **L** of the original model **M** [e.g., *Lyster et al.* 1997].

$$\mathbf{B}_{t+\Delta t} = \mathbf{L}\mathbf{B}_t^a\mathbf{L}^{\mathrm{T}} + \mathbf{Q} \tag{7}$$

where matrix **L** is defined as follows:

$$\mathbf{L} = \frac{d\mathbf{x}_{t+\Delta t}}{d\mathbf{x}_t} . \tag{8}$$

Matrix **Q** is the error covariance matrix representing errors added to vector x during model integration resulting from the model being imperfect.

The dimension of vector **x** as well as matrices **B** and **L** can be rather large. The version of ROSE model used in this work contains 19 vertical levels and 1152 (36x32) grid points at each level. This results in a total of N_x=21,888 grid points. Significant computational difficulties arise when dealing with vectors and matrices of such high dimensions. Approximately 3 5 GB (N_v x N_x x 8 bytes) is required to store either **B** or **L** matrix and implementation of direct full matrix multiplication (as in equation (7)) is clearly impossible on most present-day computers.

Certain simplifications are required in order to make the analysis feasible Following *Menard et al.* [2000] and *Menard and Chang* [2000], we compute the diagonal elements of **B** using (5) and parameterize the off-diagonal elements as follows

$$b_{ij} = \sqrt{b_{ii}b_{jj}}\exp\left(-\frac{\Delta r_{xy}^2}{2L_{xy}^2}\right)\exp\left(-\frac{\Delta r_z^2}{2L_z^2}\right) \tag{9}$$

where Δr_{xy} and Δr_z represent horizontal and vertical distances between locations i and j, and L_{xy} and L_z are adjustable parameters. Thus only N_v real numbers representing the diagonal elements of **B** have to be stored, and the off-diagonal elements can be computed as needed using (9).

Several different forms of the off-diagonal element parameterization were suggested [e.g., *Menard et al.* 2000] and *Menard and Chang* [2000]. We found that our results depend mostly on the values of the correlation length L_{xy} and L_z rather then on the functional form of the parameterization

When parameterization (9) is introduced, expression (7) describing time evolution of the error covariance is replaced, following *Menard et al.* [2000] and *Menard and Chang* [2000], by the following equations for evolution of diagonal elements of **B**, or variances b_{ii}:

$$b_{ii}(t+\Delta t) = \tilde{b}_{ii}(t+\Delta t) + q_{ii}(t) \tag{10}$$

$$\tilde{b}_{ii}(t+\Delta t) = \mathbf{M}(b_{ii}(t)) \tag{11}$$

$$q_{ii}(t) = \left[\varepsilon\ x_i(t+\Delta t)\cdot\Delta t\right]^2 . \tag{12}$$

The off-diagonal elements then are computed from the updated diagonal elements using (9).

Term $b_{ii}(t+\Delta t)$ in (10) represents the updated variance. To compute the updated variance, we assume that variance b of x is modified during the model integration in the same way as x itself This assumption is justified provided that the model is only weakly nonlinear. Thus, to compute the updated variance at time $t+\Delta t$, we simply integrate the model forward using previously obtained variances $b_{ii}(t)$ as initial conditions. Equation (11) thus symbolically represents transformation of variances by the model. The assumption of weak nonlinearity is discussed further in Summary and Discussion.

Term $q_{ii}(t)$ represents additional error introduced to account for imperfections of the model. As seen from (12), it is

proportional to the modeled field itself and the length of the integration period Δt. The proportionality coefficient ε is a tunable parameter and is discussed further in the next section. Observational error covariance **O** is assumed to be diagonal, with its diagonal elements set to the observational error variance, as described in section 3. Finally, the represen-tativeness error covariance matrix **R** is also assumed to be diagonal, with its elements computed as

$$r_{ii} = (ry_i)^2$$

where r, the relative representativeness error, is an adjustable parameter discussed in more detail in the next section. One should keep in mind that the assumption of diagonal covariance matrices is, in general, incorrect and can lead to additional errors.

The length of the assimilation window is 1 hour, which is twice the value of the model time step. During this time, MLS observes approximately 55 ozone profiles. Since, as described in section 3, seven pressure levels are used per profile, the size of vector **y** is approximately 385, and the matrix inversion required in (5) can be performed in a reasonable time.

5. Adjustable Parameters

Parameters L_{xy}, L_z, ε, and r need to be tuned to achieve best agreement with observations. The adjustment is performed using two criteria, the χ^2 (chi-square) diagnostics described by *Menard et al.* [2000] and *Menard and Chang* [2000] and analysis of differences between forecast and observations (the so-called OmF analysis).

For each assimilation analysis the scalar quantity χ^2 is computed as

$$\chi^2 = \left(\mathbf{y}-\mathbf{H}(\mathbf{x})\right)^{\mathrm{T}} \left(\mathbf{HBH}^{\mathrm{T}} + \mathbf{O}+\mathbf{R}\right)^{-1}\left(\mathbf{y}-\mathbf{H}(\mathbf{x})\right). \quad (13)$$

Note that, as explained by *Menard et al.* [2000], and *Menard and Chang* [2000]

$$\left\langle \left(\mathbf{y}-\mathbf{H}(\mathbf{x})\right)\left(\mathbf{y}-\mathbf{H}(\mathbf{x})\right)^{\mathrm{T}} \right\rangle \quad (14)$$

is the error covariance of the OmF residuals, or $\mathbf{y}-\mathbf{H}(\mathbf{x})$, computed from the actual observations and the forecast, while

$$\left(\mathbf{HBH}^{\mathrm{T}} + \mathbf{O}+\mathbf{R}\right) \quad (15)$$

is the OmF error covariance estimated in the course of the assimilation analysis. Angular brackets denote arithmetic means. Therefore we expect that

$$\left\langle \chi^2 \right\rangle \equiv N. \quad (16)$$

where N is the number of observations used in the analysis. Thus χ^2 diagnostics allows one to examine consistency of the error covariance parameterization. If the average value of χ^2/N is less than 1, than either estimated background error or estimated representativeness error is larger than their actual values computed according to (14). Note that if the assimilation is performed long enough and sufficient amount of observations are introduced into the system, the estimated background error covariance **B** does not depend anymore on its initial value. Instead, its value is completely determined by the tunable parameters. We followed these rules for determining optimal values of ε and r: (1) Run assimilation system and examine χ^2. If the average value of χ^2 increases (decreases) with time, increase (decrease) ε (2) Repeat until the average value of χ^2 does not show a trend. (3) If the average value of χ^2/N is larger (smaller) than 1, increase (decrease) r.

By following these rules we, in effect, systematically compare forecasted concentrations with observations that were not yet used in the assimilation. On the basis of results of such comparisons we tune the error growth parameter ε and the representativeness error. Assuming that observations are only weakly correlated in time, the validity of the assimilation system is thus systematically validated through comparisons with independent data.

It is difficult to perform rigorous optimization following these rules due to rather high computational requirements involved in running the assimilation system for prolonged periods of time. Nevertheless, as shown in Figure 1, values of $\varepsilon = 0.0135$ per hour and $r = 0.1$ give rise to rather satisfactory χ^2.

It turned out that value of χ^2 is relatively insensitive to variations of parameters L_{xy} and L_z. To tune these parameters, we performed several assimilation runs and examined the OmF residuals, $\mathbf{y}-\mathbf{H}(\mathbf{x})$. We varied L_{xy} from 500 to 5000 km and L_z from 0.2 to 1.5 of the standard atmospheric scale height. Best results were achieved for L_{xy} = 1000 km and L_z=0.4 of the standard atmospheric scale height. Cross sections of the resulting monthly and zonally averaged OmF residuals corresponding to these parameter values are discussed further in the next section.

6. Results

The assimilation scheme described above was used to assimilate UARS MLS ozone observations for all of 1993. To spin-up the system and prepare appropriate initial conditions, we first run the system for November and December of 1992 and assimilated all MLS observations available for that period. As observed by *Levelt et al.* [1998], assimilation changes model results dramatically. Differences between "pure" model simulations and the assimilation analysis that we observed are of the same order as those shown by *Levelt et al.* [1998]. After the initialization period, the system was run for the 12 months of 1993 with assimilation of all available MLS data. In this section we will analyze results of assimilation for January 1993. Results of the assimilation for all months can be found at http://acd.ucar.edu/~boris/research.htm.

An example of analyzed instantaneous global stratospheric ozone distribution constrained by MLS ozone data and the estimated analysis errors are shown in Plate 1 for January 13, 1993. Owing to specifics of the UARS observational pattern, there were no measurements north of approximately 35°N for several weeks prior to this date. As a result, one can see rather large variances around the North Pole and fairly small variances south of 35°N. Note a tongue of tropical, ozone-rich air extending into the Northern Hemisphere. The density of observations in the tropical region is fairly high, and the associated variance is therefore low. As seen in Plate 1, this intrusion significantly lowers variance at higher latitudes. The model thus helps to constrain analysis in the nonobserved regions by retaining and "advecting" information in the system.

Plate 2 shows time evolution of the analyzed ozone and its variance at 10 mbar and 0° longitude as a function of latitude. The data are plotted every hour. At the beginning of the month, UARS was observing mostly the Northern Hemisphere, from approximately 35° S to about 80° N. As can be seen from the plot, the corresponding variance increases with time in the region

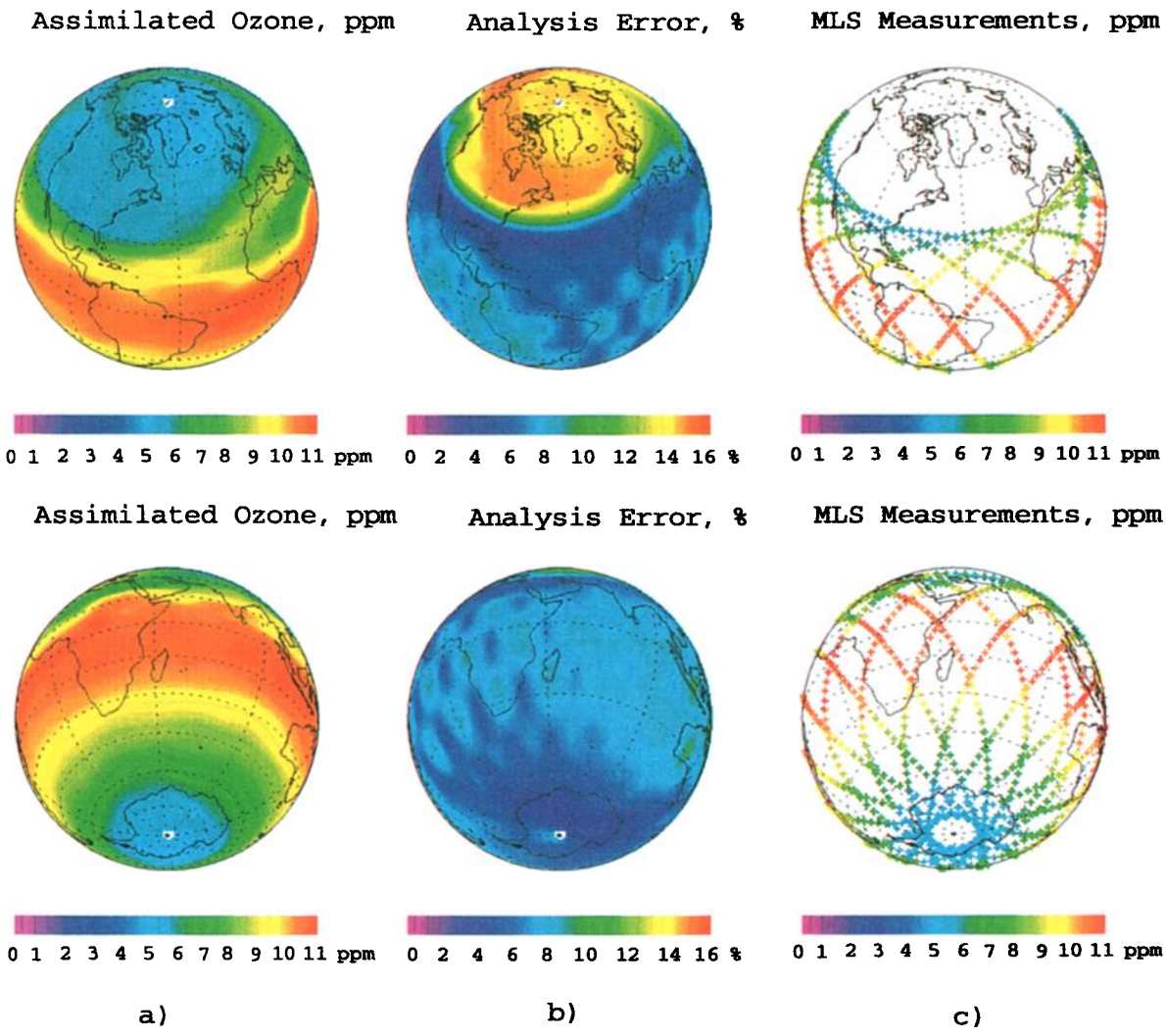

Plate 1. Results of the assimilation for January 13, 1993, at 10 mbar. (a) Analyzed ozone at 1300 UT, ppmv; (b) analysis error in percent at 1300 UT; (c) MLS ozone observations.

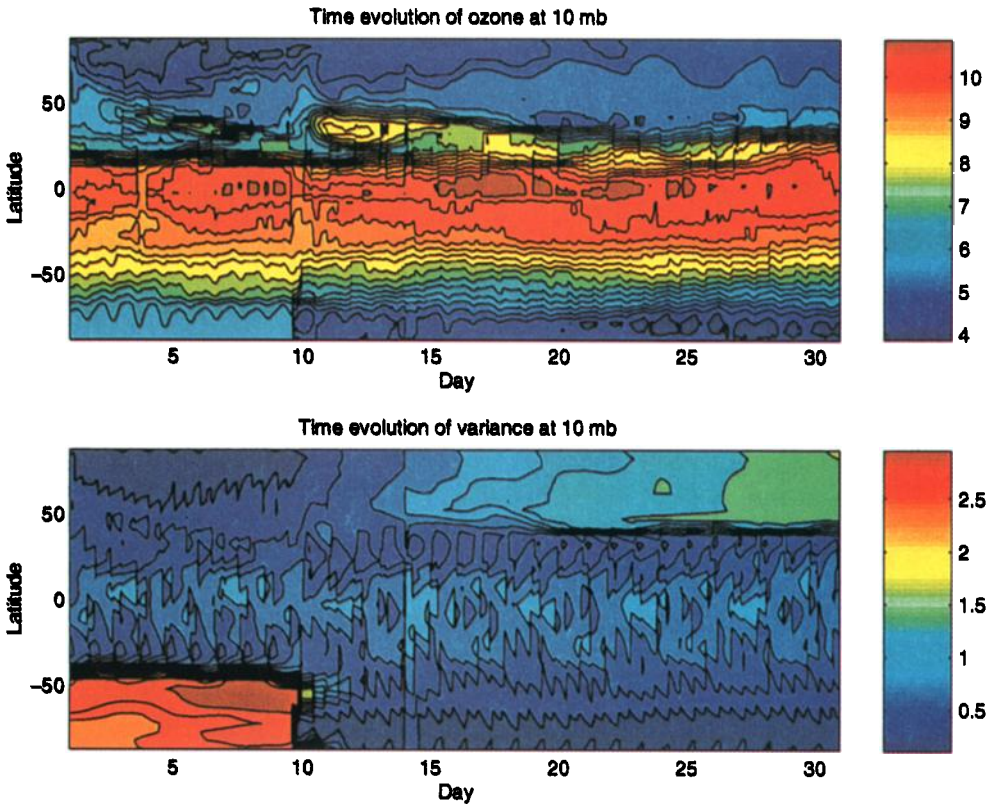

Plate 2. Time evolution of ozone (ppmv) and square root of the variance (ppmv) at 10 mbar and 0° longitude for January 1993.

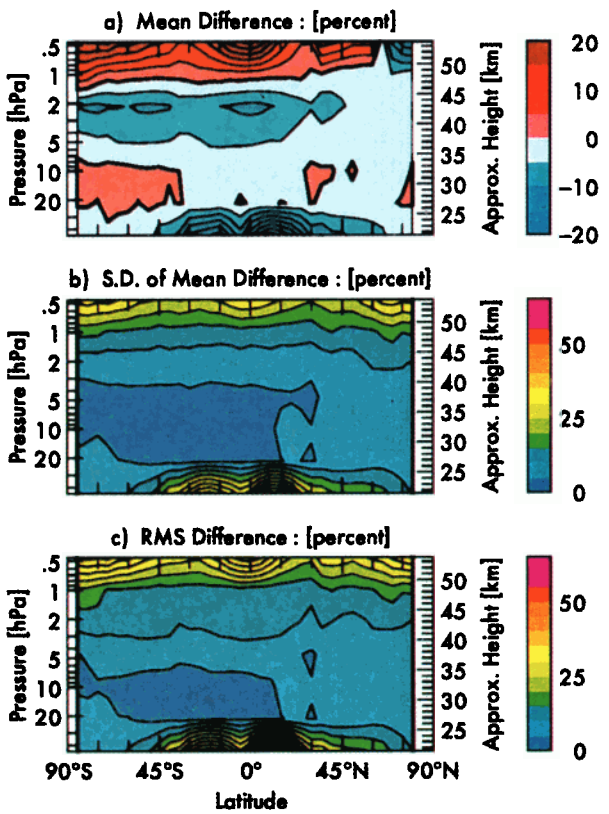

Plate 3. Monthly averaged for January 1993 cross sections of (a) mean difference or bias $\langle y\text{-}H(x)\rangle$; (b) standard deviation from the mean $\sqrt{\langle (y\text{-}H(x)\text{-}\langle y\text{-}H(x)\rangle)^2\rangle}$; and (c) RMS differences of OmF $\sqrt{\langle (y\text{-}H(x))^2\rangle}$.

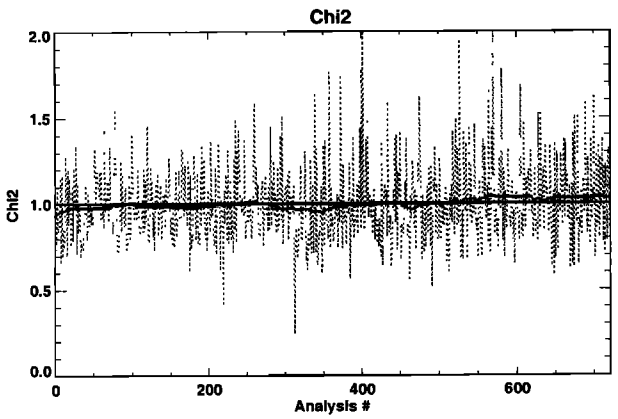

Figure 1. Plot of χ^2/N as a function of time for January 1993. Each value is computed for one assimilation analysis (1 hour in length, containing approximately 55 profiles or 365 points). The thick line is the 100-point (~4 days) moving average.

where no MLS observations are available, while it stays approximately constant in the region where there are observations. Around January 10 the satellite switched to the south looking mode, and the analyzed ozone and variance fields at high southern latitudes change accordingly, while the variance starts to grow at high northern latitudes.

Figure 2 presents individual ozone profiles measured by another UARS instrument, Halogen Occultation Experiment (HALOE), and the corresponding profiles from the assimilation system together with estimated errors of the assimilation results. Note that HALOE measurements were not used in the assimilation. HALOE ozone measurements have errors of the order of 5% in the region between 25 and 60 km according to the HALOE Data Quality Document http://haloedata.larc.nasa.gov/Haloe/qualityvalidationstatus.html. The assimilation results are given at every model pressure level. Note that the error bars are generally smaller at the pressure levels where MLS observations are present than on the pressure levels that are not directly constrained with observations. Overall, the agreement between our results and independent HALOE measurements is fairly good at least at the MLS observed levels, and HALOE observations are practically always located within the estimated error bars of the assimilation analysis. The upper three plots in this figure showing rather large analysis errors on January 2, 5, and 9 correspond to high southern latitudes where no MLS data were available for about 30 days. The agreement between HALOE data and the analysis is nevertheless reasonable. It appears than that the assimilation scheme overestimates the analysis variance when there are long temporal gaps in observations. We believe that this is due to the linear error growth assumption as discussed in the Summary and Conclusions section. At the 6.8 mbar level the agreement between HALOE and MLS is generally much worse than at the adjacent levels. This is because there is no retrieved MLS ozone at this level in MLS version 4 data. We expect this problem to disappear when the more recent MLS version 5 data are used.

To validate the estimated variances of our results, we computed root-mean-squared (RMS) differences between the assimilation analysis and all available HALOE ozone measurements. Comparisons between the calculated RMS differences and average estimated analysis variances are shown in Figure 3 for several 10°-wide latitudinal bands. Corresponding zonally and time-averaged HALOE and assimilation ozone profiles are also shown. In most cases a reasonable agreement exists between RMS differences of our results and HALOE measurements and the independently estimated analysis errors. In those cases when the two estimates differ considerably, calculated assimilation variances are usually larger than the RMS values, implying that our error estimates are conservative.

A somewhat more systematic view of differences between MLS observations and our results is shown in Plate 3. This plot presents monthly and zonally averaged root-mean-squared differences between the analysis and observations, corresponding bias, and standard deviation from the mean. As one can see, both the bias and the standard deviation are within 5-10% in the region between approximately 25 and 45 km. This is to be expected, as both the accuracy and precision of the MLS data decrease beyond this region. While OmF residuals can be considered a by-product of the assimilation, as seen from Plate 3 their analysis provides an important quality control mechanism and highlights systematic differences between the analysis and data.

7. Summary and Discussion

Development of data assimilation techniques for use with global atmospheric chemistry transport models can be a very resource-intensive task. At the same time, use of such techniques for analysis of chemical observations of atmospheric composition promises significant benefits and is becoming increasingly important as more space-based instruments are deployed. These instruments supply vast amounts of information that need to be analyzed in a systematic and rigorous manner. Properly designed data assimilation methodologies will allow scientists to better understand and interpret chemical composition and dynamical processes of the Earth's atmosphere. These tools also provide means to systematically evaluate and validate the associated numerical models.

One of the main goals in developing the assimilation scheme was to provide a simple to use and computationally fast tool for use by scientists using and developing three-dimensional chemistry transport models. At the present time the developed technique has been used in a number of different CTMs for assimilation of several types of observations. While results of this and several other research efforts [*Lamarque et al.*, 1999; *Levelt et al.*, 1998; Collins et al., submitted manuscript, 2000] demonstrate the usefulness of the proposed approach, there is certainly room for improvement. Several assumptions made in the formulation of the data assimilation method can be discarded in the future implementations.

One of the most important simplifications is the assumption of the weakly nonlinear model. While this assumption appears to be justified for ozone in the lower and middle stratosphere, in the upper stratosphere the ozone photochemical lifetime becomes short due to increasing importance of several chemical processes. To compute time evolution of the error covariance under these conditions using the extended Kalman filter, one needs to know the linearization of the model [*Khattatov et al.*, 1999]. While such computations are feasible for simple zero-dimensional box models as described by *Khattatov et al.* [1999], applications to 3-D models appear to be practically impossible at the present time.

We have assumed that values of adjustable parameters L_{xy}, L_z, ε, and r in (10), (11), and (12) are independent of time and

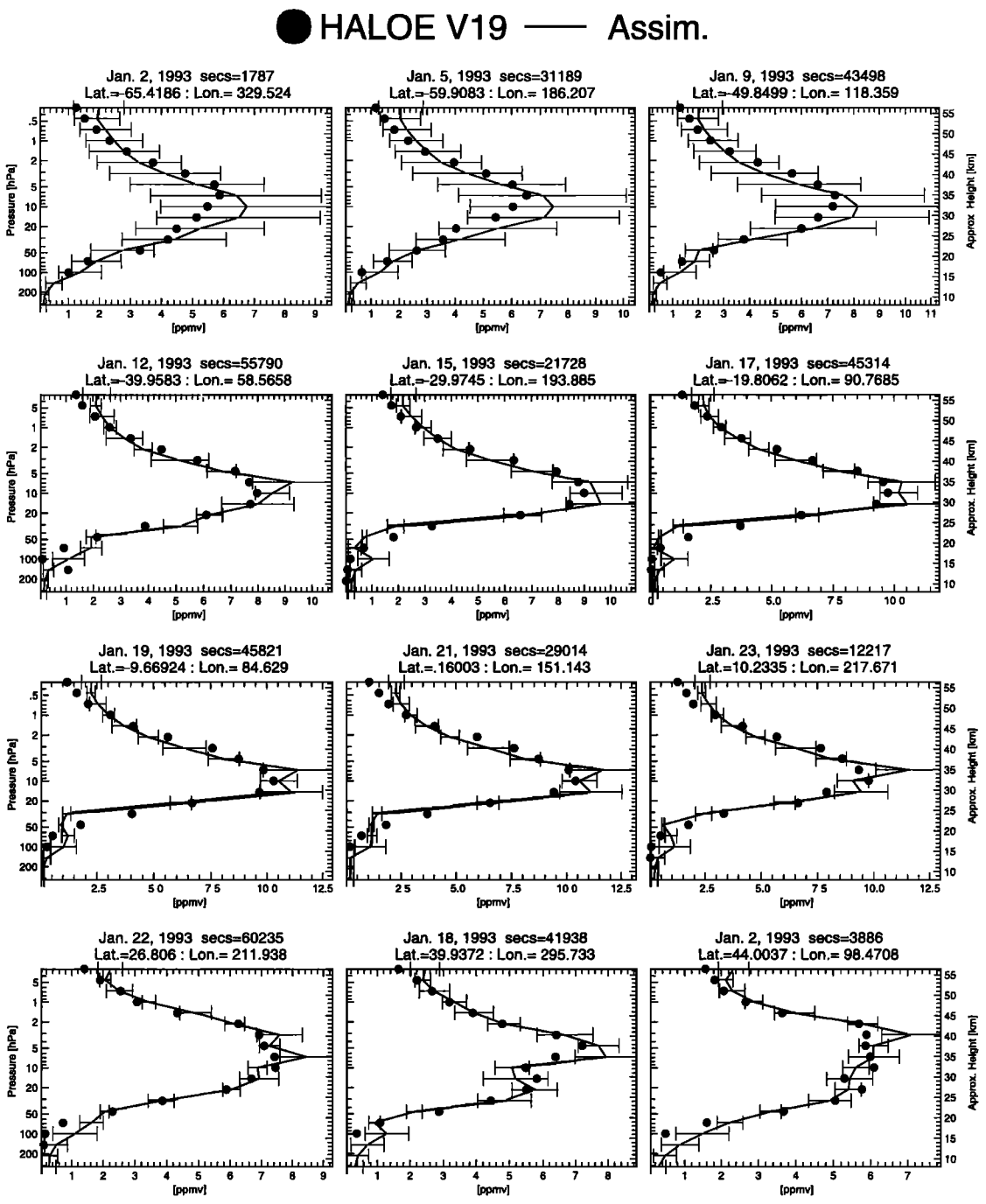

Figure 2. Individual ozone observations from HALOE (solid circles), ozone profiles from the assimilation system (solid line), and the corresponding error bars of the assimilation at several different locations for January 1993. See Results section for discussion of the assimilation analysis error bars.

geographical location. While this assumption significantly simplifies tuning these parameters, it is certainly not necessary and can be changed. For instance, the vertical correlation length is expected to be larger in the stratosphere than in the troposphere. The model error growth rate will be larger in the regions where dynamical processes are not modeled very well (e.g., tropics) or where some chemical processes are omitted from the model formulation. Accordingly, χ^2 analysis should be performed at several different geographical regions and pressure levels, rather than globally as was done in this work.

It is reasonable to assume that the correlation length in the zonal direction is larger than that in the meridional direction. Thus (9) can be amended to include one more exponential term and thus another free parameter. This, of course, will make the process of tuning the scheme even more cumbersome.

Finally, the assumption of linear growth of the errors during model integration is unrealistic. It is reasonable to conclude that modern CTMs simulate distributions of most atmospheric chemicals with errors less than 100-200%. If the linear error growth assumption were correct, the analysis error would not

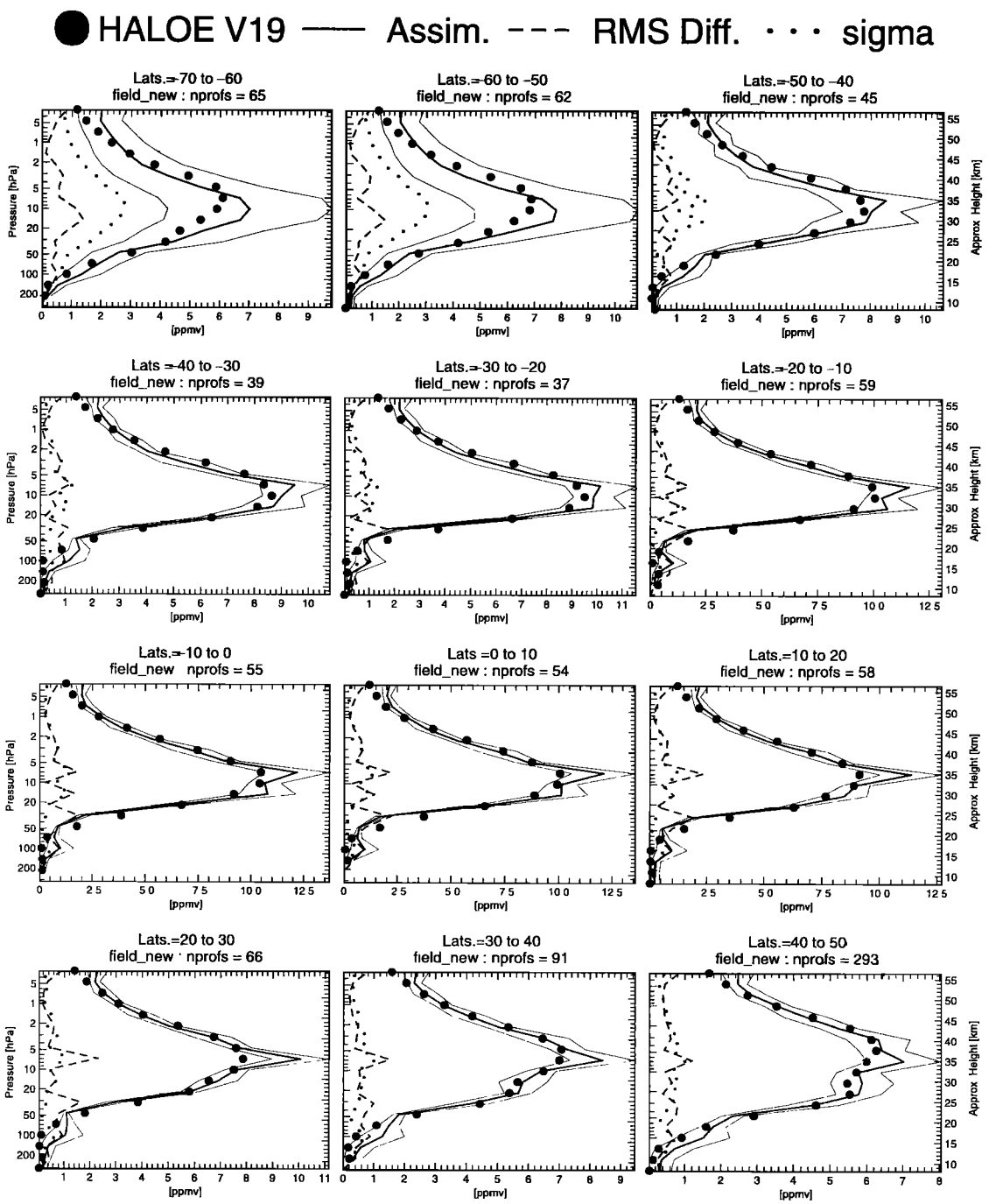

Figure 3. Monthly (January) and zonally averaged ozone profiles from HALOE (solid circles) and the assimilation (thick lines) for different latitude bins. Also shown are the average RMS differences between HALOE and assimilation (dotted lines) and average assimilation analysis error (dashed lines). Thin solid lines represent error bars of the assimilation analysis.

have an upper bound at all in the regions where no observations are assimilated for prolonged periods of time This explains the large error bars seen at the upper three plots in Figure 2.

Ultimately, soundness of the adopted simplifications can be assessed by comparing assimilation results with independent data, such as measurements not yet used in the assimilation (Plate 3) or observations from different instrument (Figures 1 and 2). As one can see from these plots, our results show reasonable agreement with both sources of independent data in the region from about 25 to 50 km.

The three-dimensional sequential data assimilation method described here allows fast and accurate mapping of sparse and scattered satellite data to regular time and space grid This process provides means to easily visualize instantaneous distributions of the observed chemical species (e.g , Plate 1) and significantly facilitates comparisons with correlative data (e g., Figure 2). Short- as well as long-term temporal variability at fixed geographic locations is readily observed (Plate 2). Estimated variance of the resulting instantaneous three-dimensional distributions supplies an important quantitative

quality indicator allowing for cross-validation of different sources of measurements. We expect that the developed technique will find numerous applications in processing and validation of data from future satellite missions, including EOS-TERRA and EOS-CHEM. The Fortran code of the assimilation scheme together with assimilation results for 12 months of 1993 can be accessed at http://acd.ucar.edu/~boris/research.htm.

Acknowledgments. The authors are grateful to Bill Randel, Steve Massie, Peter Lyster, and Henk Eskes for fruitful discussion and valuable comments We thank two anonymous reviewers for their detailed comments and help in improving the manuscript. The authors acknowledge use of the UKMO wind and temperature data in this research and thank UKMO for providing these data. Research of B.V.K. was supported by the NASA UARS guest investigator project under NRA-97-MTPE-04 (contract S10109-X) and the NASA MOPITT project. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

References

DeMore, W. B., et al., Chemical kinetic and photochemical data for use in stratospheric modeling: Evaluation 11 of the NASA Panel for Data Evaluation, *JPL Publ.*, 1994.

Elbern, H., H. Schmidt, and A. Ebel, Variational data assimilation for tropospheric chemistry modeling, *J. Geophys. Res., 102,* 15,967-15,985, 1997

Eskes, H. J., A. J. M. Piters, P. F Levelt, M. A. Allaart, and H. M. Kelder, Variational assimilation of total ozone satellite data in a 2D lat-lon tracer-transport model, *J. Atmos. Sci.*, in press, 2000.

Fisher, M., and D. J. Lary, Lagrangian four dimensional variational data assimilation of chemical species, *Q. J. R. Meteorol. Soc., 121,* 1681-1704, 1995.

Froidevaux L. et al , Validation of UARS Microwave Limb Sounder ozone measurements, *J. Geophys. Res., 101,* 10017-10060, 1996.

Granier, C., and G. Brasseur, Ozone and other trace gases in the Arctic and Antarctic regions: Three-dimensional model simulations, *J. Geophys. Res., 96,* 2995-3011, 1991.

Khattatov, B. V., J. C. Gille, L. V. Lyjak, G. P. Brasseur, V. L. Dvortsov, A. E. Roche, and J. Waters, Assimilation of photochemically active species and a case analysis of UARS data, *J Geophys Res., 104,* 18,715-18,737, 1999.

Lamarque, J.-F , B. V. Khattatov, and J. C. Gille, Assimilation of Measurement of Air Pollution From Space (MAPS) CO in a global three-dimensional model, *J. Geophys. Res., 104,* 26,209-26,218, 1999.

Levelt, P. F., B. V. Khattatov, J. C. Gille, G. P. Brasseur, X X. Tie, and J. Waters, Assimilation of MLS ozone measurements in the global three-dimensional chemistry-transport model ROSE, *Geophys. Res. Lett., 25,* 4493-4496, 1998.

Lorenc, A. C., Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc., 112,* 1177-1194, 1986.

Lyster, P. M., S. E. Cohn, R. Menard, L.-P. Chang, S.-J. Lin, and R. Olsen, An implementation of a two-dimensional filter for atmospheric chemical constituent assimilation on massively parallel computers, *Mon. Weather. Rev., 125,* 1674-1686, 1997.

Menard, R., and L.-P. Chang, Stratospheric assimilation of chemical tracer observations using a Kalman filter, part II,: Chi-square validated results and analysis of variance and correlation dynamics, *Mon. Weather Rev., 128,* 2672-2686, 2000.

Menard, R., S.E. Cohn, L.-P. Chang, and P.M. Lyster, Stratospheric assimilation of chemical tracer observations using a Kalman filter, part I, Formulation, *Mon. Weather Rev., 128,* 2654-2671, 2000a.

Reber, C. A., et al., The Upper Atmosphere Research Satellite (UARS) mission, *J. Geophys. Res., 98,* 10,643-10,647, 1993.

Rnshojgaard, L. P., I. Stajner, and G.-P. Lou, The GEOS Ozone Data Assimilation System, *Adv. Space Res., 25,* 1063-1072, 2000.

Rose, K., and G. Brasseur, A three-dimensional model of chemically active trace species in the middle atmosphere during disturbed winter conditions, *J. Geophys. Res., 94,* 16,387-16,403, 1989

Smolarkiewicz, P.K., and P.J Rasch, Monotone advection on the sphere: An Eulerian versus semi-Langrangian approach, *J. Atmos Sci., 48,* 793-810, 1991.

Swinbank, R., and A. O'Neill, A stratosphere-troposphere data assimilation system, *Mon. Weather Rev., 122,* 686-702, 1994.

G. P. Brasseur, Max Plank Institute for Meteorology, Hamburg, Germany, D-20146

J. C. Gille, B. V. Khattatov, J. F. Lamarque, L. V. Lyjak, and X. X. Tie, National Center for Atmospheric Research, P. O. Box 3000, Boulder, CO 80307. (boris@ucar.edu)

P. F. Levelt, Royal Dutch Meteorological Institute, De Bilt, Netherlands, 3730.

R. Menard, Air Quality Research Branch, Meteorological Service of Canada, 2121, Transcanada Highway, 5th Floor, Dorval, Quebec, Canada, H9P 1J3.

(Received May 19, 2000; revised July 25, 2000;
accepted August 1, 2000.)